# Homework 3

**Huy Quang Lai**
Department of Computer Science and Engineering
Texas A&M University
College Station, Texas 77843
`lai.huy@tamu.edu`

## 1 Prediction

### 1.1 Train Ridge Regression

To tune the hyper-parameters of the Ridge Regression model, a grid search is performed over different values of `alpha`, stored in `lambda_vals`. The algorithm splits the data into training and validation sets and trains the model on the training set for each value of `alpha`. Then the algorithm evaluates the model on the validation set and chooses the value of `alpha` that gives the highest ROC-AUC score.

The algorithm stores the ROC-AUC score of the model for each value of `alpha` and iteration in a dictionary `aucs`. After all iterations are completed, it calculates the mean ROC-AUC score of the model for each value of `alpha`. The optimal value of `alpha` would be the one that gives the highest mean ROC-AUC score. When running this method, the optimal value is 10 with a ROC-AUC score of 0.9751.

### 1.2 Train Lasso

The function `train_lasso` performs hyper-parameter tuning for the Lasso regression model using similar approach to the Ridge Regression hyper-parameter tuning. Tuning is doing through a grid search approach. The algorithm loops through a set of lambda values, defined in `lambda_vals`, and trains a Lasso model for each lambda value for `n` iterations. In each iteration, the algorithm fits the Lasso model on the training data and calculates the AUC-ROC score on the test data. The AUC-ROC score for each lambda value is stored in a dictionary `aucs` as a list of dictionaries where each dictionary contains a single lambda value and its corresponding AUC-ROC score.

After the loop, the function calculates the mean AUC-ROC score for each lambda value by computing the mean of all AUC-ROC scores across all iterations. It then stores these mean AUC-ROC scores in a dictionary `lasso_mean_auc` and prints them out. Finally, the function returns the dictionary `lasso_mean_auc`.

To determine the optimal lambda value, the algorithm selects the lambda value that gives the highest AUC-ROC score, and uses that as the optimal lambda value. Upon executing the code, the lambda value of 0.1 was found with a AUC-ROC score of 0.9707.
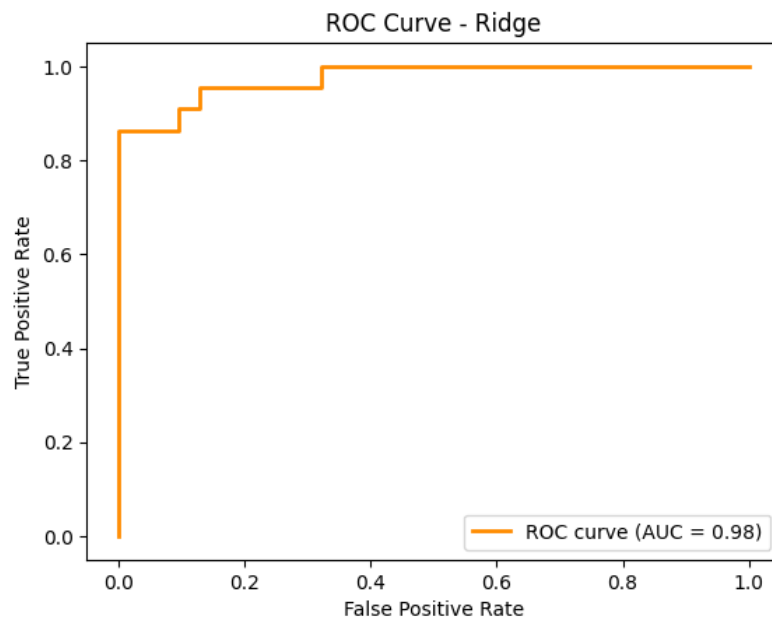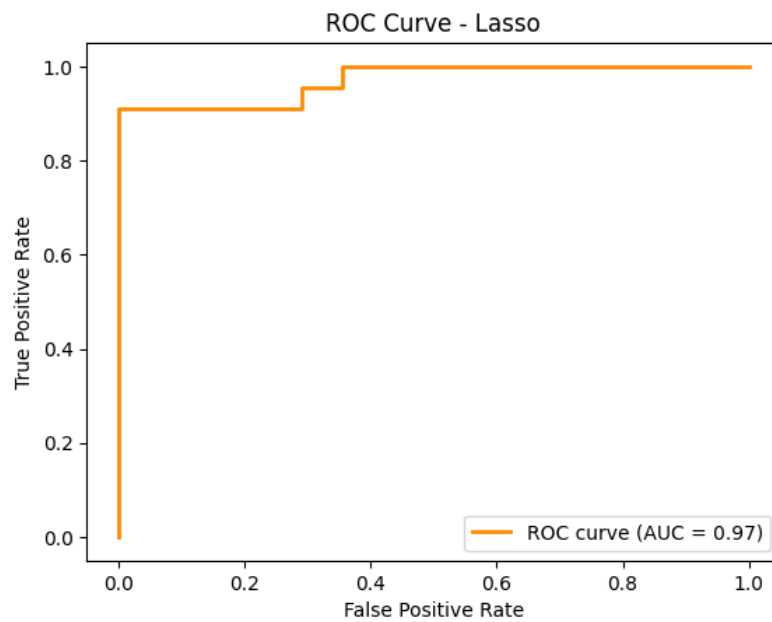
Figure 1: ROC Curve for Ridge Regression



Figure 2: ROC Curve for Lasso Regression

# References

[1] Watt, Jeremy, Borhani, Reza  & Katsaggelos, Aggelos Konstantinos (2016) Machine Learning Refined.

[2] Konasani, Venkata Reddy  & Shailendra Kadre (2021) Machine Learning and Deep Learning Using Python and TensorFlow.